Geophys. J. Int. (1990) **100,** 19–32

Single-link cluster analysis as a method to evaluate spatial and temporal properties of earthquake catalogues

Cliff Frohlich[1] and Scott D. Davis[1,2]

[1]*Institute for Geophysics, University of Texas at Austin,* 8701 *N. Mopac Boulevard, Austin, Texas* 78759-8345, *USA*
[2]*Department of Geological Sciences University of Texas at Austin P.O. Box* 7909 *Austin, Texas* 78713-7909, *USA*

Accepted 1989 June 9. Received 1989 June 9; in original form 1988 November 2

SUMMARY

Single-link cluster analysis is a straightforward method to quantitatively measure the degree of clustering or isolation of groups of elements in a set, such as a catalogue of earthquakes. To apply single-link cluster analysis to a set of N earthquakes, individual earthquakes are first linked to their nearest neighbours to form event sub-groups. The process is then repeated and each sub-group linked to its nearest neighbour, recursively, until N-1 links are found to join all the earthquakes. This paper shows how knowledge of these links can be used to divide a group of earthquakes into any number of spatial clusters. With various modifications this analysis can identify earthquake nests, isolated events, aftershock sequences, and zones of seismic quiescence. To illustrate this method we apply it to; (i) a global data set of 2178 earthquakes having m_b of 5.8 or greater reported by the International Seismological Centre (ISC) between 1964 and February 1986, and (ii) sets of earthquakes having m_b of 4.9 and greater as reported by the ISC, occurring in Central America and in the Aleutians. To facilitate comparison with real data, this study also investigates the distribution of link lengths for synthetic events placed randomly on 1-D (circular line), 2-D (circular area) and 3-D (spherical volume) geometries. If we remove links between nuclear explosions from the ISC data, the resulting link distribution is quite similar to that for synthetic events placed along a 1-D fault. For the shortest links, the similarity improves markedly if the synthetic fault has a finite width, or if relative location errors cause events to be located away from the fault trace, with a standard deviation of about 15 km.

Key words: cluster analysis, earthquakes.

INTRODUCTION

The distribution of earthquakes in both space and time is clearly non-random, as exemplified by the existence of earthquake aftershocks, swarms and nests. However, in some sense the distribution in space is less random than the distribution in time. While tests to detect time clustering often compare sequences of events to random, or Poisson processes (e.g. Vere-Jones 1970; Shlien & Toksoz 1970; Utsu 1972; Frohlich & Davis 1985; Frohlich 1987), these are seldom or never used to evaluate the spatial relationships among events. Methods to evaluate spatial clustering of earthquakes include measuring the distribution of nearest-neighbour distances (Suzuki & Suzuki 1965), analysis of frequency of events in equal area meshes (Suzuki & Suzuki 1966), analysis of second- and higher-order moments of various catalogues (Kagan 1981a,b; Kagan & Knopoff 1981; Reasenberg 1985; Eneva & Pavlis 1988), and the determination of parameters (usually depending on magnitude) for various models of spatial and temporal influence (Kagan & Knopoff 1976, 1978; Prozorov & Dziewonski 1982).

The present paper describes a completely different method, single-link cluster analysis (SLC), to evaluate spatial clustering. This paper and our previous preliminary research (Frohlich & Davis 1986; Davis & Frohlich 1986; Wardlaw, Frohlich & Davis 1988) are apparently the first application of SLC to sequences of earthquakes. This is surprising, as there is a relatively large amount of literature on SLC (e.g. Ling 1973; Sneath & Sokal 1973; Hartigan 1975; Ripley 1984; Day & Edelsbrunner 1984), and it has been used to evaluate clustering in numerous other fields including taxonomy, social science, astronomy, seismic interpretation (Amanzade & Chatterje 1984), and linguistics.

SLC has several features which make it an attractive

method to apply to earthquake sequences. Its hierarchical nature provides a natural scaling which can define clusters or find isolated events at both the most global and most local levels. It is algebraically and computationally appealing (see Appendix A), in that it naturally expresses the relationship between individual events as well as between different groups of events. Finally, the observed tendency in SLC for clusters to form linear chains (e.g. Sneath & Sokal 1973) seems quite appropriate for earthquakes, which often occur along linear features such as faults or spreading ridges.

The intent of the present paper is to show how the framework of SLC provides a natural way to quantitatively describe certain well-known characteristics of earthquakes. These include the isolated character of intraplate events, the tendency of events to form linear groups along fault lines, and the existence of earthquake 'nests'. We shall also speculate on how SLC could be used to quantitatively identify aftershocks and zones of seismic quiescence. The present paper will not undertake an exhaustive comparison of various clustering methods applied to earthquakes.

SINGLE-LINK CLUSTERING AND DEFINITION OF A d-CLUSTER

We say a group of earthquakes forms a *d-km cluster* (Fig. 1) if each earthquake is within d or fewer km of at least one other earthquake within the group. Thus, when two earthquakes in a d-km cluster are separated by more than d km, there must exist a sequence of earthquakes in the group connecting the two, such that each event in the sequence is within d or fewer km of the next event.

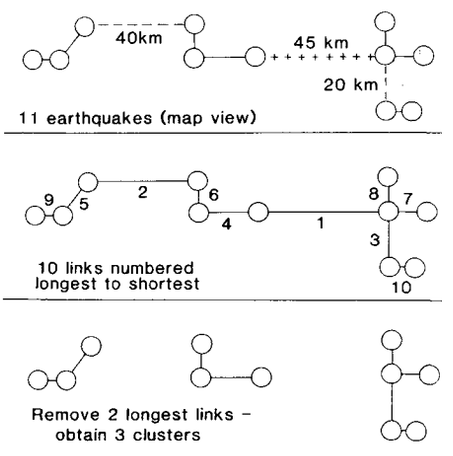

Figure 1. Explanation of linkage and d-km clustering in single-link cluster analysis (SLC). Let each of the 11 earthquakes (circles) shown in map view be linked to its nearest neighbour (solid lines in top diagram), then let each group formed in this way be linked to its nearest neighbouring group, etc., (dashed and + + + lines) until all 11 events are joined by 10 links. If the two links larger than 20 km are removed (links 1 and 2 in middle diagram), the remaining linked groups form three 20-km clusters (bottom diagram). Within each 20-km cluster all events are within 20 km of another event in the cluster, and no events are within 20 km of any event in the other clusters. In general, for N events there will be $N-1$ links, and $k+1$ clusters will be formed when the k longest links are removed. If d_k and d_{k+1} are the lengths of the kth and $(k+1)$th longest links, then the corresponding $k+1$ clusters are d-clusters for all d such that $d_k > d > d_{k+1}$.

In Fig. 1 (top), the 11 events shown form three 20-km clusters since all the events in each cluster are joined by links of length 20 km or shorter (solid lines and 20 km dashed line). Since the individual clusters are 40 and 45 km apart, respectively, (40 km dashed lines and 45 km + lines), these three clusters remain distinct unless d is equal to or larger than 40 km, when the two clusters on the left join together to form a cluster. When d is larger than 45 km, all three clusters merge and the entire group of 11 events forms a single 45-km cluster.

For any particular subset G of events forming a d-cluster, three convenient statistics to describe the clustering are: (1) the number n of events in G, (2) *the clustering distance* d_C, the minimum value of d for which the events in G are linked together as a d-cluster, and (3) the *isolation distance* d_I which is the maximum value of d for which the events in G form a d-cluster, without absorbing events outside G. Clearly d_C is a measure of how close the events are to one another, whereas d_I is a measure of how isolated they are from other groups of events. G will be visibly recognized as a distinct cluster or nest if n is large and d_C is small. A single event or cluster is perceived as isolated when $d_I \gg d_C$. For example, in Fig. 1, d_C is 20 km for the rightmost group of five events because they form a d-cluster when d is 20 km, but not for smaller values of d. They join the middle group when d is 45 km or larger, so d_I is 45 km.

The clustering process defined in this way is called *single-link* clustering because two clusters merge when they are joined by the shortest link joining them. SLC is said to be hierarchical (Fig. 1). For example, if no two earthquakes coincide, for very small d the N events will form N distinct d-clusters, each consisting of a single, isolated event. As d increases, individual events are linked to one another, so that after k links are formed, there remain $N-k$ clusters. Finally, for some large value of d all N events are linked by $N-1$ links into a single, large cluster. Thus as the $N-1$ links are added one at a time, from the smallest to the largest, they define a natural hierarchy of clustering for the N events into $N-1$ clusters, then $N-2$ clusters $\cdots$ then 2 clusters, then a single cluster. In fact, in SLC an ordered list of linked events, or a map showing the events joined by the $N-1$ links, provides a complete description of the clustering process. However, for many purposes it is often useful to consider only the lengths of the $N-1$ links.

For computation, the SLC problem simplifies considerably (see Appendix A and Fig. 1) because links joining nearest neighbours will always be among the $N-1$ SLC links, and because the 'trees' of events formed by nearest-neighbour chains form a higher order set of 'events'. These trees can in turn be analysed to find the nearest-neighbour trees, recursively, until all the SLC links are formed.

The mathematics of the SLC problem depends critically on the definition of the concept of distance between elements in a set. For example, in taxonomy the elements are plant or animal species, and the problem of the taxonomist is to define a distance between species in such a way as to produce meaningful results from the SLC analysis. In this paper, the elements are points in space, the earthquake hypocentres, and except where explicitly stated otherwise, the distance metric is the usual 3-D Euclidean distance.

RESULTS FOR LARGE EARTHQUAKES IN THE ISC CATALOGUE

Description of linkage

The first group of earthquakes studied with SLC analysis was all events reported by the International Seismological Centre (ISC) with body wave magnitude m_b of 5.8 or higher. This group included 2178 events occurring between 1964 January and 1986 February. Thus the SLC analysis determined 2177 links (Fig. 2), with lengths ranging from 0.3 to 3465 km (Table 1).

In general, links longer than about 1000 km join peculiar, isolated events or event groups to large groupings of earthquakes which make up major tectonic structural groups. For example, the longest link, link 1, joins a group of earthquakes in Hawaii to earthquakes offshore of the western United States. Links 2, 5 and 9 join nuclear explosions in Novaya Zemyla to earthquakes in Asia and earthquakes near Iceland. Link 3 joins earthquakes in the eastern Pacific, while link 6 connects earthquakes in the southern Atlantic. Link 4 is the largest link of a chain of earthquakes joining an earthquake in eastern Canada to earthquakes in the Caribbean. Links 7 and 8 join earthquakes in Africa. Links 19 and 23 join earthquakes in Australia to earthquakes in Java.

Removal of the 38 links longer than 1000 km leaves 39 1000-km clusters. The largest cluster, comprising the majority of the events, contains 1735 circum-Pacific earthquakes, linked from California through Alaska, Japan, and the western Pacific, including Tonga, New Zealand and portions of central Asia. Central America, South America and the South Sandwich arc form the next largest cluster. Many of the remaining clusters are single (e.g. Table 2) isolated earthquakes (16 clusters), or small groups of two, three or four earthquakes (eight clusters). The majority of clusters occur in ridge or fracture-zone areas where long chains of earthquakes break into clusters because the density of earthquakes having $m_b \geq 5.8$ is low. Nuclear explosions in Novaya Zemyla (22 events) and Tahiti (seven events) form two of the clusters.

There are 248 links with lengths between 200 and 1000 km. These links join major groups to other major groups, or join sub-groups within major groupings of events. The removal of these links separates the circum-Pacific 1000-km cluster, described above, into numerous smaller clusters. Thus Alaskan earthquakes are now distinct from earthquakes in the 48 contiguous states, and earthquakes in

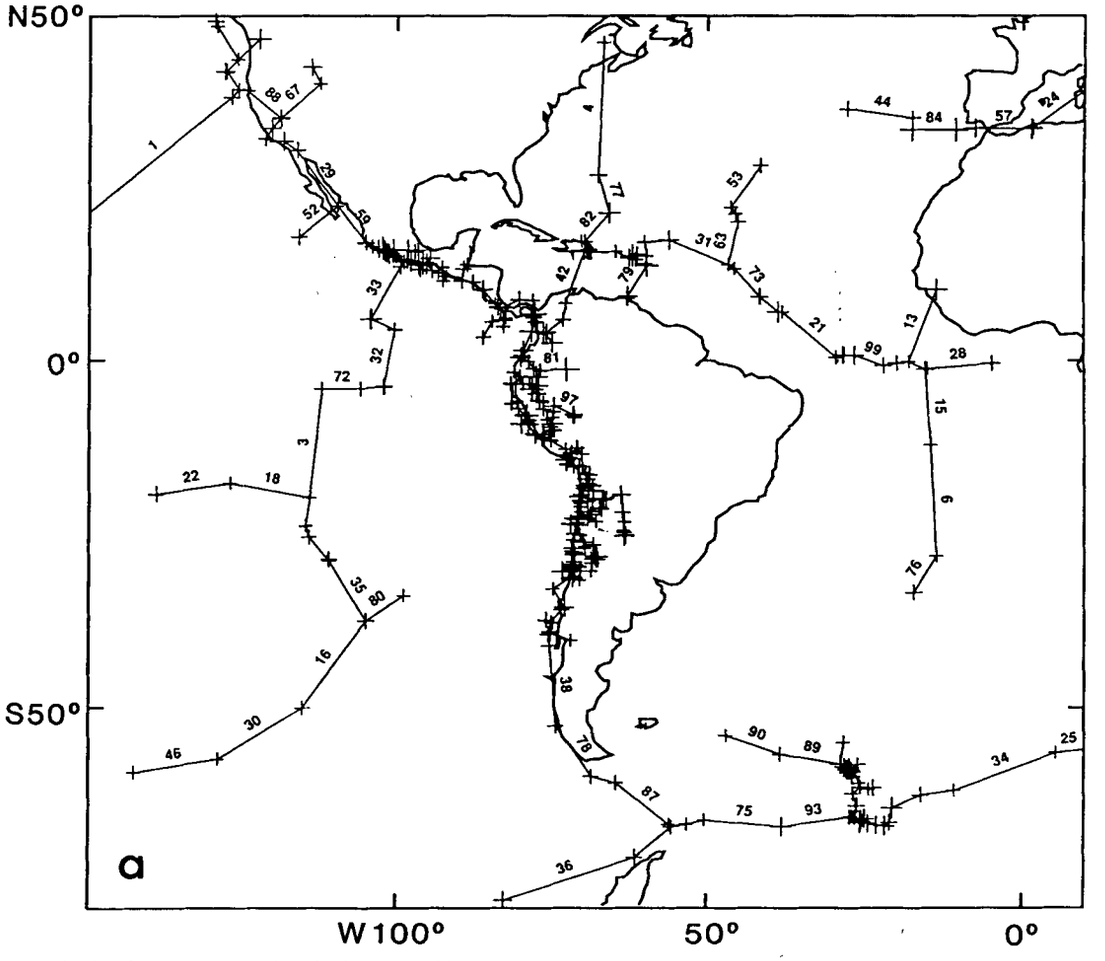

Figure 2. Global distribution of earthquakes (+) reported by the ISC between 1964 January and 1986 February with m_b of 5.8 or greater, and linkage determined by SLC. Maps are Mercator projections for (a) western hemisphere; (b) eastern hemisphere; and a polar projection for (c) north polar region. The numbers shown above some links in the Mercator maps label the 100 longest links, ranked in order of decreasing size, and the 200 longest links in the polar projection.

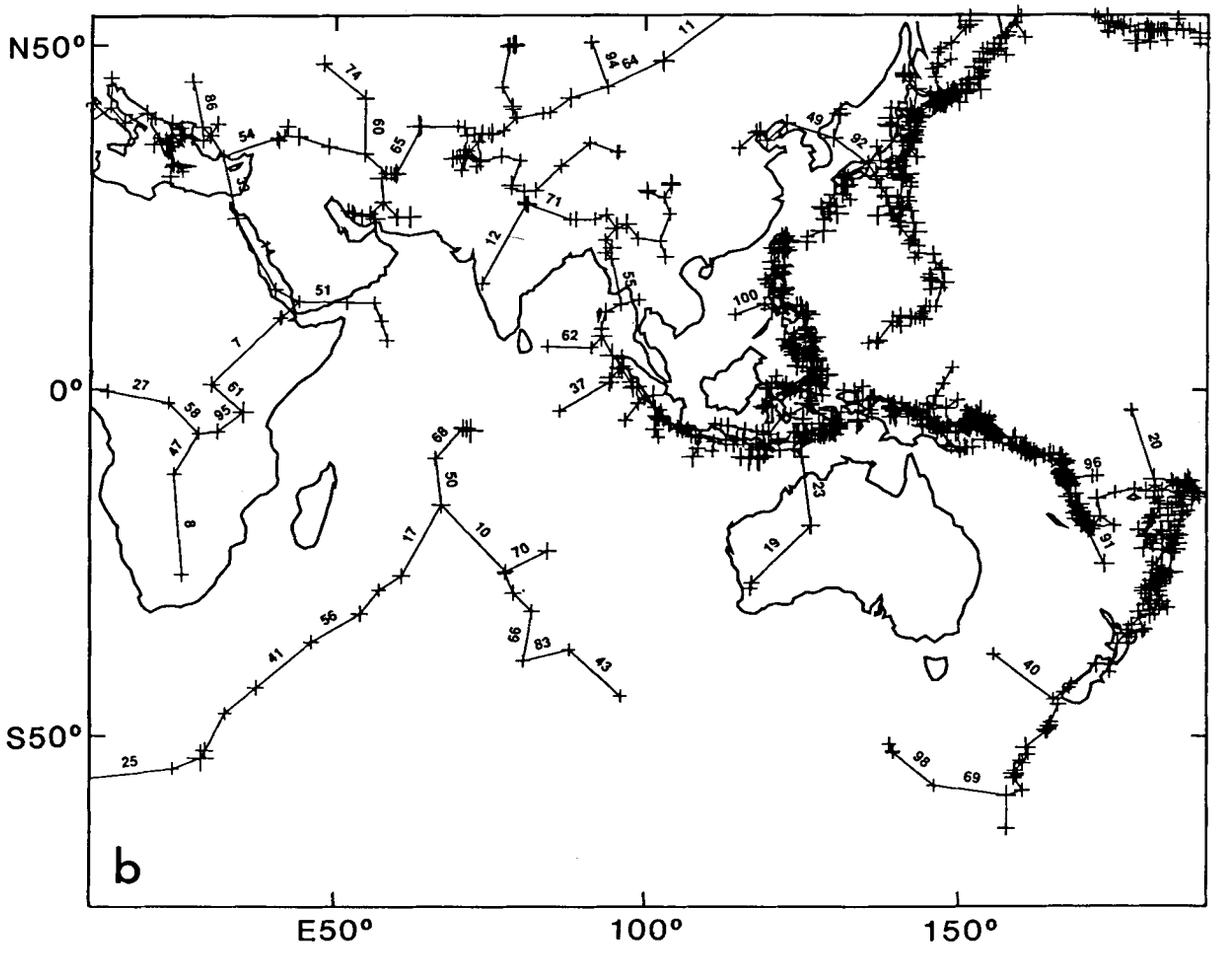

Figure 2b. (*continued*)

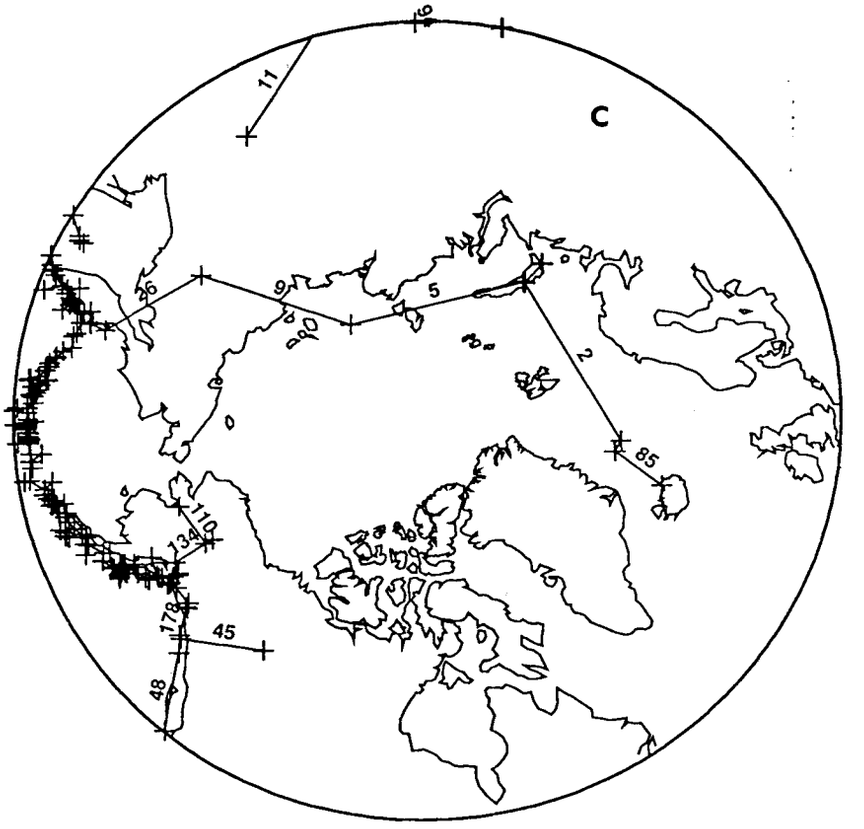

Figure 2c. (*continued*)

Table 1. Tectonic interpretation of link lengths.

Link Length (km)	Number of Links	Type of Groups Connected	Typical Example
> 1000	38	isolated events to other isolated events, or to major tectonic groups	Hawaii to North America; earthquakes in Africa to one another
200 - 1000	248	major groups to major groups, events in sparse fault chains to one another	South America to South Sandwich; Tonga-Kermadec to New Hebrides
30 - 200	1206	clusters of earthquakes within major groups to similar clusters	majority of earthquakes in Tonga joined
5 - 30	588	events in "nests;" aftershock sequences	intermediate Hindu -Kush earthquakes, many 1964 Alaskan aftershocks
< 5	97	nuclear explosions	Novaya Zemyla

Table 2. Especially isolated earthquakes in the set of 2178 ISC events having $m_b > 5.8$, and occurring between 1964 January and 1986 February. The Table includes all events which are (1200-km, 2) isolated, i.e. when all links up to length 1200 km are included, the listed events are in a group of 2 or fewer linked events.

	origin	time	latitude	longitude	depth	d_I [a] (km)	link #[b]	n[c]	Region
1.	Jan. 9, 1982	12:53	46.99° N	66.65° W	10	1913	4	1	New Brunswick
2.	June 14, 1977	21:39	14.02° S	14.49° W	21	1868	6	1	South Atlantic Ridge
3.	June 3, 1981	05:47	36.00° S	17.08° W	10	1868	6	2	South Atlantic Ridge
	Nov. 22, 1984	00:50.	30.90° S	13.53° W	10				
4.	July 1, 1976	11:24	29.51° S	25.17° E	29	1718	8	1	Union of South Africa
5.	Aug. 25, 1964	13:47	78.15° N	126.65° E	34	1701	9	1	East of Severnaya Zemlya
6.	Jan. 18, 1967	05:34	56.68°N	120.95° E	5	1528	11	1	Eastern Russia
7.	Dec. 10, 1967	22:51	17.54° N	73.84° E	27	1509	12	1	India
8.	Dec. 22, 1983	04:11	11.85° N	13.51° W	11	1411	13	1	Northwest Africa
9.	Jul. 31, 1983	10:26	20.13° S	126.94° W	10	1358	18	1	South Pacific Ocean
10.	Oct. 14, 1968	02:58	31.54° S	117.00° E	18	1340	19	2	Western Australia
	June 2, 1979	09:47	30.73° S	117.21° E	6				
11.	March 24,1970	10:35	22.08° S	126.65° E	3	1226	23	1	Western Australia

[a] d_I is the distance between event group and nearest neighbouring event group.
[b] link # is rank of link joining event group to nearest neighbouring group, with 1 being longest link, 2 next longest, etc.
[c] n is the number of events in cluster.

Mexico, Central America, northern South America and southern South America are all in separate clusters. Tonga, the Kermadecs, the Marianas, the Philippines and the South Sandwich arcs are now all distinct, densely populated, 200-km clusters. The deep earthquakes in South America now form clusters separate from the intermediate earthquakes, and the long chain-like groups along ridges and fracture zones have almost entirely disappeared, splitting into numerous isolated clusters of one, or a few, earthquakes.

The majority of links (1206) have lengths between 30 and 200 km. These links join earthquakes within individual tectonic groups, such as the South Sandwich arc (Fig. 3) or the Tonga–Kermadec arc (Fig. 4). When they are removed, the remaining 30-km clusters having more than a few earthquakes tend to form nests. For example, for the 40 earthquakes in the South Sandwich region, all but one of the links shorter than 30 km join earthquakes in the intermediate-depth nest having 10 members in the north of the arc (A in Fig. 3) or the shallow nest of five earthquakes (B in Fig. 3) in the south. For the 155 earthquakes in the Tonga–Kermadec region, there are shallow 30-km clusters having 17 members in the south (B in Fig. 4) and 12 members in the north (A in Fig. 4). The next four largest 30-km clusters are shallow clusters of six and four earthquakes, respectively, and two clusters of four deep earthquakes. For these earthquakes, which all have

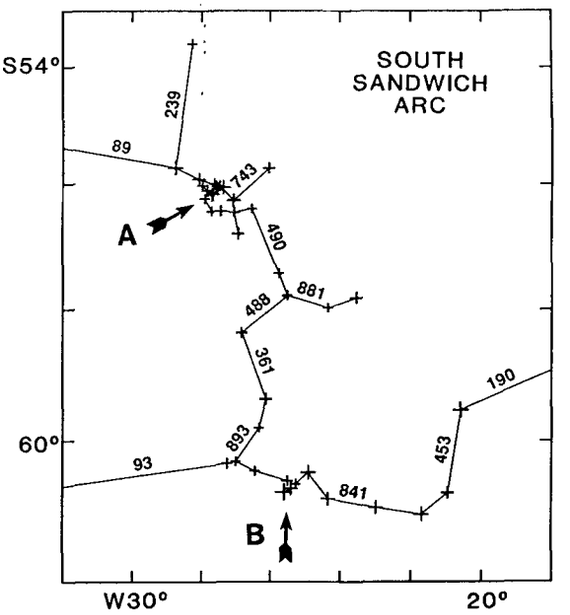

Figure 3. Linkage determined by SLC for earthquakes with m_b of 5.8 or greater in the South Sandwich arc region. The numbers label the 900 longest links in the global catalogue. Note that the South Sandwich group becomes a single cluster with the removal of links 89, 93, and 190, and it splits into two sub-clusters upon the removal of link 361. 'A' and 'B' denote clusters described in the text.

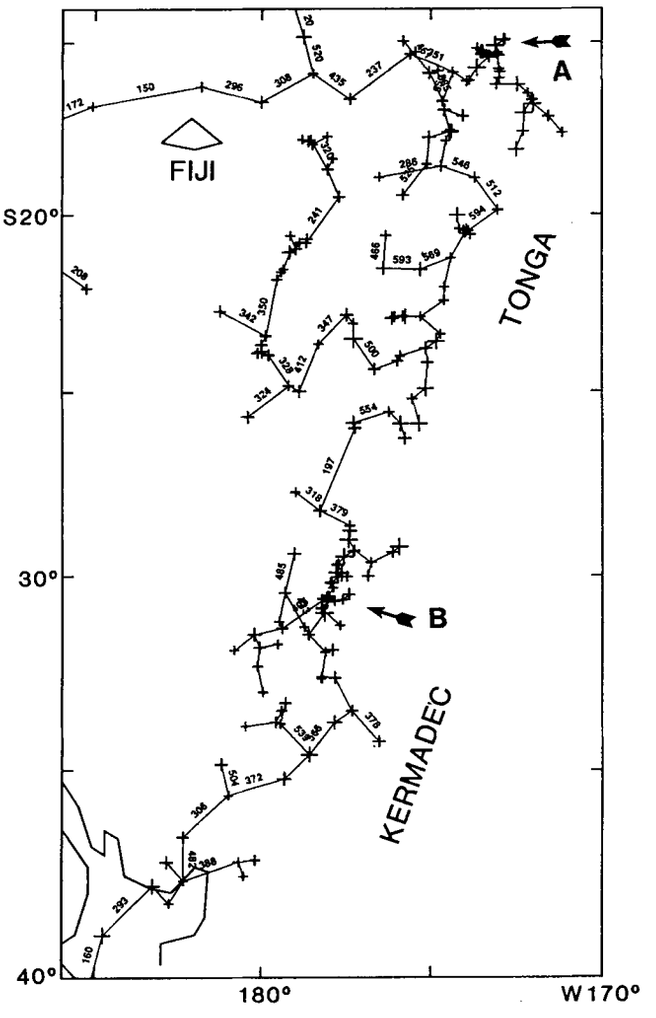

Figure 4. Linkage determined by SLC for earthquakes with m_b of 5.8 or greater in the Tonga–Kermadec region. The numbers label the longest 600 links. Note that the Tonga–Kermadec group becomes a single cluster with the removal of links 150 and 160, and that Tonga and Kermadec split into separate sub-clusters with the removal of link 197. 'A' and 'B' denote clusters described in the text.

Table 3. Highly clustered event groups in the set of 2178 ISC events having $m_b > 5.8$, and occurring between 1964 January and 1986 February. The table includes all events which are (50-km, 20) clustered, i.e., when all links up to length 50 km are included, the listed events are in a group of 20 or more linked events.

Region	mean latitude	mean longitude	mean depth	n_{50}[a]	n_{80}[b]	notes
1. New Ireland-Solomons	6.00° S	154.41° E	54	68	123	A
2. Kurile Islands	43.81° N	147.90° E	63	81		
3. Eastern Kazakh	49.90° N	78.83° E	0	61	61	B
4. New Britain	5.66° S	151.75° E	46	31	123	A
5. Solomon Islands	10.17° S	160.91° E	56	28	36	
6. Kermadec Islands	30.07° S	177.76° W	33	27	38	
7. Molucca Passage	1.38° N	126.31° E	33	25	30	C
8. Rat Islands	51.36° N	161.72° W	42	21	24	

[a] - n_{50} is number of events in cluster joined by 50 km links

[b] - n_{80} is number of events in cluster joined by 80 km links

A - Clusters 1. and 4. joined by 59 km link

B - Nuclear explosions - 59 of these events form (10-km, 59) cluster

C - 24 of these events form (32-km, 24) cluster

magnitude of 5.8 and larger, the more intense nests (Table 3) occur in subduction zones, and have mean depths of 30–60 km.

In addition to forming nests, many links in 30-km clusters are apparently due to aftershock activity. Indeed, of the 588 links with lengths between 5 and 30 km, 103 join earthquakes occurring within 10 days, or less, of one another. How does this compare to what would be expected if links joined earthquakes having random times generated by a Poisson process acting over a time interval T (here 22.16 yr)? We consider a link with one earthquake occurring at time t within the time interval $(0, T)$, and determine the probability f that the other earthquake will occur within the interval $(t - \Delta, t + \Delta]$. The expected fraction will be $2\Delta/T$ if t is in the sub-interval $(\Delta, T - \Delta)$, which has length $T - 2\Delta$. If t is in the interval $(0, \Delta)$, the expected fraction will be $(t + \Delta)/T$, which has a mean value of $3\Delta/2T$ over the entire sub-interval of length Δ. The fraction is also $3\Delta/2T$ for t within the sub-interval $(T - \Delta, T)$. Thus, straightforward algebra shows that the weighted average for the fraction f over the entire interval $(0, T)$ is

$$f = \frac{\Delta}{T}\left(2 - \frac{\Delta}{T}\right).$$

For 588 random events 1.5 links would be expected to join earthquakes occurring within 10 days of one another, while 103 such links for the 588 earthquakes having lengths between 5 and 30 km are observed. In contrast, for the 657 links longer than 100 km, exactly two occur in the natural earthquake data whereas 1.6 would be expected for random events. In general, the data suggest that aftershock activity affects linkage for links as long as about 100 km, but will not affect linkage for longer links.

For the 97 links shorter than 5 km, 79 join nests of nuclear explosions in the Soviet Union or in Tahiti. In addition, seven links join earthquakes occurring less than 10 days apart, whereas the formula above suggests that fewer than one would be expected in the absence of aftershock activity.

Distribution of link lengths

SLC provides a framework for arbitrarily dividing our data into any number of clusters, or defining d-clusters for any arbitrarily chosen value of d. However, is there any particular number of clusters, or particular value of d which is 'natural' for our data? If clumping or natural clustering did occur in a data set it might be expected that the distribution of link lengths would exhibit a distinct change near the particular value of d which characterized the clustering. No such gross change is visible in the distribution for the ISC earthquake data (Fig. 5). For links with lengths between about 20 and 200 km, the distribution is very nearly a straight line, with no jumps or changes in slope.

How similar is the distribution of links joining earthquakes to links between events randomly distributed along a line or over a surface (see Appendix B)? 2178 synthetic events were placed randomly along the equator of a sphere (1-D distribution) and along the surface of a sphere (2-D distribution) in order to investigate this. For initial comparison to the earthquake data, we chose spheres with

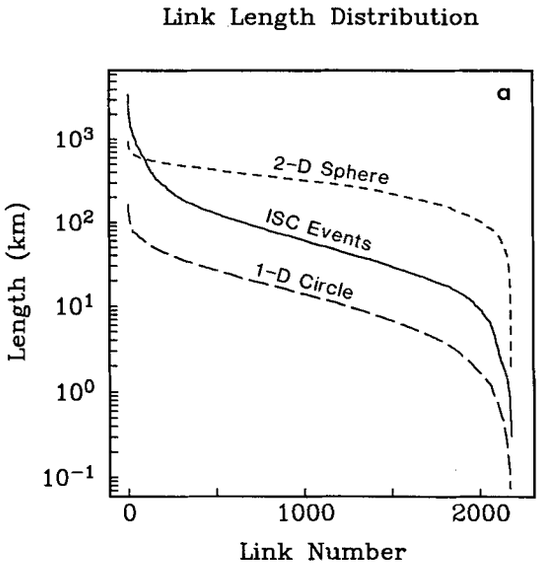

Figure 5a. Cumulative distribution of link lengths for the events shown in Fig. 2 (solid line) together with the distribution of link lengths for 2178 events placed randomly on a circle (1-D) and on the surface of a sphere (2-D). The circle and the sphere both have radii equal to that of the Earth.

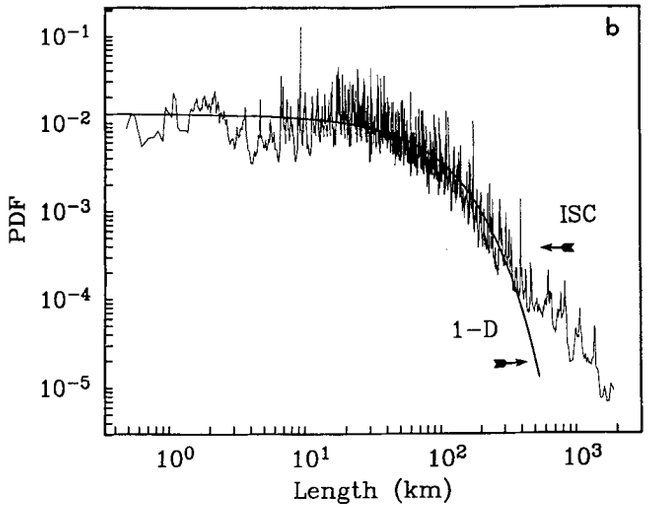

Figure 5b. Probability density function (PDF) for the lengths of links for events in Fig. 2 (thin straight line), and for the best-fitting 1-D distribution (thick line). For the best fitting 1-D distribution, 2178 events are distributed randomly along a circle with circumference 194 437 km.

radius equal to that of the Earth. We then applied SLC in each case to determine the 2177 links joining the synthetic events.

The results (Fig. 5) indicate that real earthquakes are most similar to the 1-D synthetic events. In particular, the slope of the 1-D distribution is quite close to that of the real earthquake data, whereas the slope of the 2-D distribution is quite different. For example, consider the difference between the ISC link lengths d_{ISC} and the 1-D and 2-D link lengths d_{1D} and d_{2D}. If we approximate this difference as a straight line then;

$$\log_{10} d_{ISC} - \log_{10} d_{1D} = a_{1D}(k/1000) + b_{1D}$$

$$\log_{10} d_{ISC} - \log_{10} d_{2D} = a_{2D}(k/1000) + b_{2D}.$$

If a least-squares method is used to fit the central 80 per cent (links $k = 219$ to $k = 1959$) of the graph (Fig. 5a), it is found that;

$$a_{1D} = 0.0216 \pm 0.0016$$

$$b_{1D} = 0.6864 \pm 0.0008 = \log_{10}(4.857)$$

$$a_{2D} = 0.350 \pm 0.0014$$

$$b_{2D} = -0.3472 \pm 0.0007$$

The difference a_{1D} in the slopes of the 1-D and ISC link lengths is about 16 times smaller than the difference a_{2D} in the 2-D and ISC lengths. The difference b_{1D} in the intercepts suggests that the length distribution would be most similar if we changed the scale of the 1-D distribution by multiplying each link by a factor of 4.857. This corresponds to placing 2178 random earthquakes on the equator of a sphere with a circumference of 194 437 km, or 4.857 that of the Earth.

The resulting similarity (Fig. 5b and Appendix B) in the distributions suggests that in some sense the link lengths of the ISC earthquakes are distributed as if the earthquakes occurred randomly along a single fault with a length of approximately 195 000 km. However, a more careful comparison (Fig. 6a) shows that for the real earthquakes both the smallest and largest links are significantly longer than for the 1-D synthetic events. For several reasons this is not surprising. The largest links include those joining isolated interplate earthquakes (e.g. Table 2), which do not occur in the 1-D synthetic data. The shortest links no longer join up long linear faults or delineate plate boundaries, but instead they link events occurring as 2- and 3-D clusters, and as such their links are longer than in the 1-D case. Moreover, because the uncertainties in relative locations are typically about 10–20 km, even if earthquakes did actually occur randomly along a linear fault, these uncertainties tend to increase the observed link lengths. After scaling eight of the smallest 1-D links are shorter than 0.3 km, the length of the smallest ISC link.

At about link 2060 in Fig. 6(a), we observe a peculiar secondary peak in the distribution of link ratios, and then a distinct minimum near link 2120. These features are artifacts caused by the inclusion of the nuclear explosions in the data, as both the peak and minimum disappear when the 93 links joining nuclear explosions are removed (Fig. 6b), but not when only 103 aftershocks (links shorter than 30 km joining earthquakes occurring less than 10 days apart) are removed. This is not surprising, as 79 of the 97 links shorter than 5 km join nuclear explosions.

After removing nuclear explosions, location errors alone could account for the distribution of the smaller links. Indeed, if synthetic links are placed along the equator of a sphere, but then a north–south normally distributed location uncertainty included, with a standard deviation of 15 km (Fig. 6c), the resulting distribution is almost indistinguishable from that of the nuke-free ISC data. This result might also occur even in the absence of location errors because earthquake activity occurred over fault zones having a finite width.

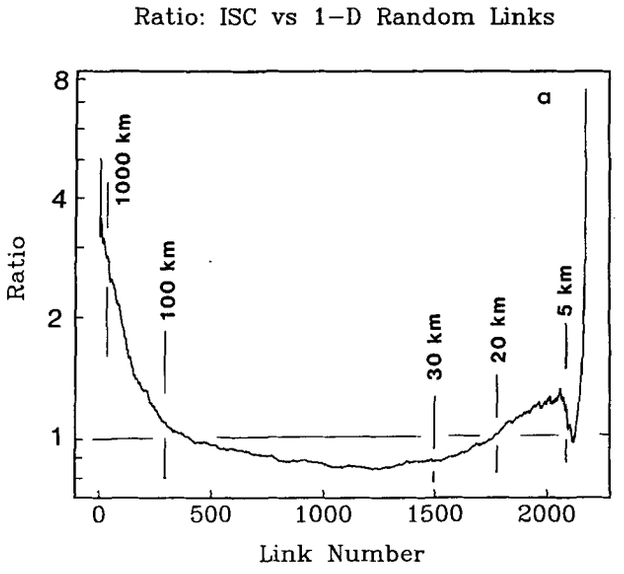

Figure 6a. Comparison of link lengths for the ISC earthquake data of Fig. 2 and random events randomly distributed along a circle with circumference of 194 437 km. The ratio plotted is the length of the kth largest link for the ISC data divided by the length of the kth largest link for the random event data. On this graph, the position of the ISC links having lengths of 1000, 200, 30, 20 and 5 km are also indicated.

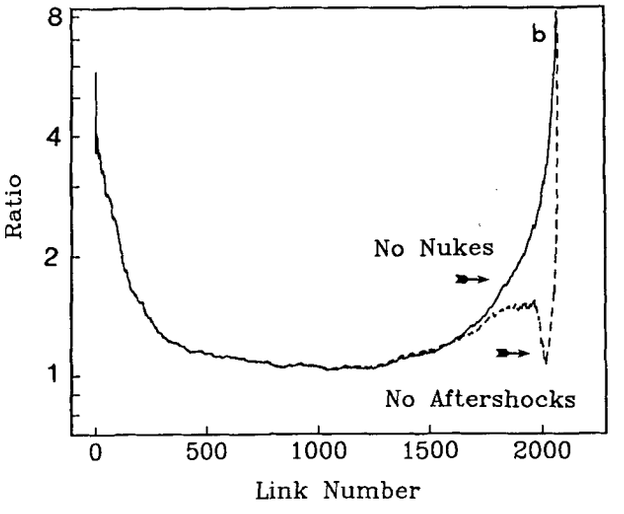

Figure 6b. Comparison of link ratios as in Fig. 6(a), except here 93 links joining nuclear explosions (solid line), or 103 links joining aftershocks (dotted line) are removed. Here aftershocks are links shorter than 30 km joining events occurring less than 10 days apart. The 1-D random link lengths in the ratio denominator are lengths expected for 2084 and 2074 links, respectively, distributed along the equator of a sphere with circumference of 163 213 km. As described in the text, this circumference is determined by a least-squares method using the data with nuclear explosives removed, for links $k = 219$ to $k = 1959$.

OTHER POSSIBLE APPLICATIONS OF SLC

Identification of seismic quiescence prior to large earthquakes

If we apply SLC to groups of earthquakes occurring in fixed time intervals in an earthquake catalogue, then the linkage

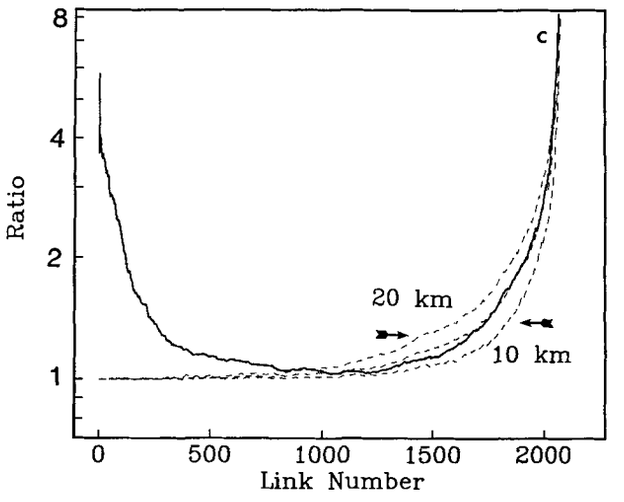

Figure 6c. Relative location errors may account for the large link ratios at small distances (high link numbers) after nuclear explosions are removed. Here the solid line is ratio as in Fig. 6(b) for the link length distribution with nuclear explosions removed. Dotted lines are link length ratios for synthetic distributions of 2084 random events distributed equatorially. However, for each pair of earthquakes there is a normally distributed relative location uncertainty of 10 km (bottom line), 15 km (middle line) or 20 km (top line). Denominator of ratio is link length distribution with nuclear explosions removed and no location uncertainty, as in Fig. 6(b).

provides an objective means of identifying time periods with unusual patterns of seismicity. In the years preceding some large earthquakes there have been reports that seismic activity in the potential rupture zone may be absent or reduced in frequency (Kelleher, Sykes & Oliver 1973; Ohtake, Matsumoto & Latham 1977; Mogi 1979; Habermann 1981). Here we shall follow Habermann, McCann & Nishenko (1983) and call this phenomenon seismic quiescence. However, there has been considerable controversy over the identification of zones of seismic quiescence, and whether such zones are at all reliable as earthquake precursors (e.g. Reasenberg & Matthews 1988; Habermann 1988).

To characterize seismic quiescence with SLC, one must specify a scale distance d_s and a time period P. A zone of seismic quiescence occurs when no links of length d_s or shorter cross a particular region during a time period of length P. This has significance only if the links do cross the region during typical intervals of length P.

For example, near Oaxaca, Mexico (Fig. 7), it was found by trial and error that a zone of seismic quiescence is evident in 1976–1977 for an interval P of 2 years and a scale distance d_s of 350 km. Following the $M_w = 7.6$ Oaxaca earthquake of 1978 November 29, no seismic quiescence is visible near Oaxaca. In Central America there seem to be several large earthquakes which are preceded by periods of quiescence (marked 'Q' in Fig. 7). However, note that regions exist which are not crossed by links (marked 'A' in Fig. 7) but which are apparently not related to the occurrence of large earthquakes. Clearly, quiescence determined by SLC analysis is meaningful for earthquake prediction only if applied to nearly complete catalogues with uniform magnitudes, attributes which may not apply to the ISC catalogue for Central American earthquakes.

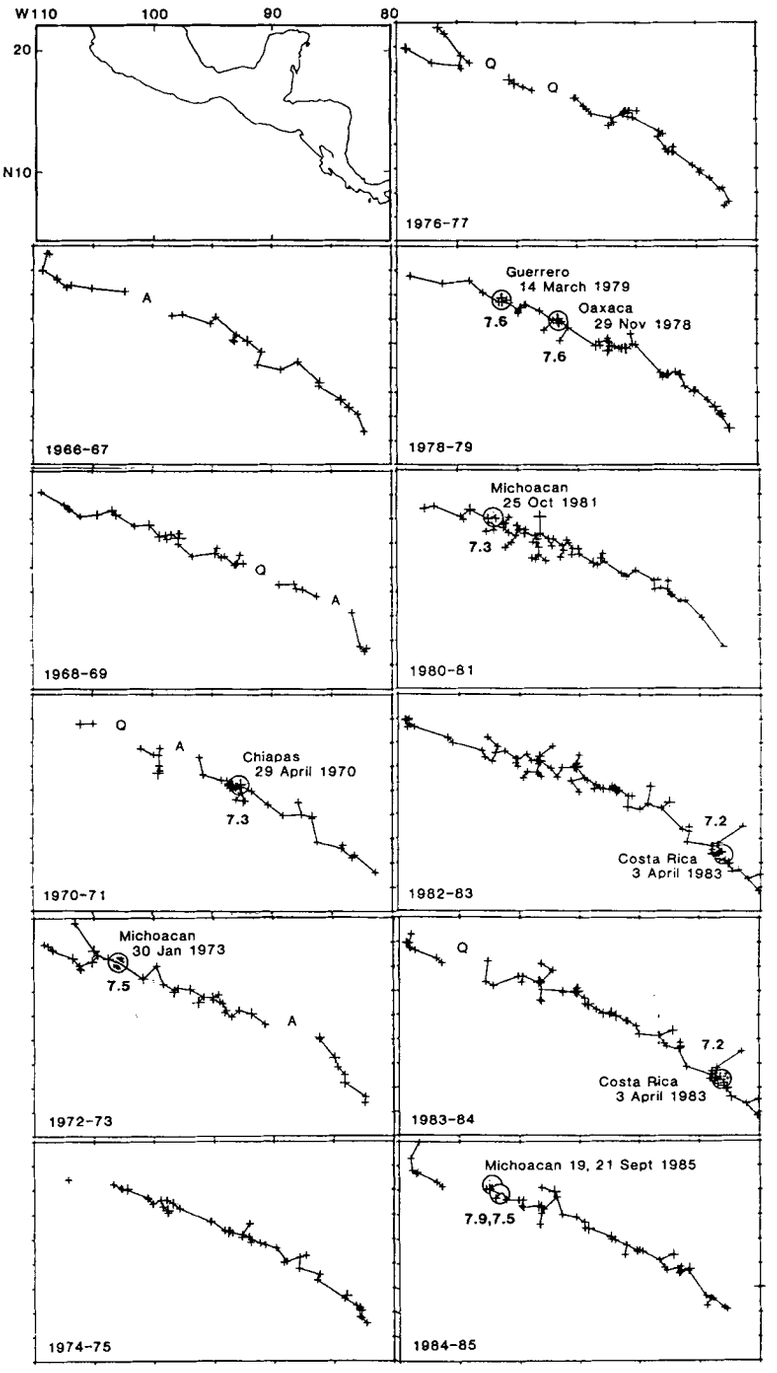

Figure 7. Pattern of 350-km clusters over 2-yr periods in the region near Oaxaca, Mexico. Top map at left shows coastlines, which are omitted in other maps because linked earthquakes approximately coincide with shoreline. The events linked by SLC are earthquakes in the ISC catalogue having m_b of 4.9 or greater occurring in each 2-yr period beginning in 1966. All links shorter than 350 km are plotted. Note the zone of seismic quiescence near Oaxaca in 1976–1977, prior to the occurrence of the 1978 November 29 earthquake. Also note other examples of quiescence prior to earthquakes with magnitudes M_S or M_W of 7.2 and greater (marked 'Q') as well as quiescence not followed by large earthquakes (marked 'A').

Nevertheless, SLC provides an objective and quantitative method for characterizing quiescence. In a more detailed ongoing study (Wardlaw, Frohlich & Davis, in preparation), we currently are evaluating how the values of P and d_s affect the expression of quiescent zones throughout the world, and how often these are associated with the occurrence of the large earthquakes.

Identification of aftershocks

As described above, among the large ISC events a significant fraction of all links shorter than about 30 km join events in aftershock sequences. However, links defined by spatial distance alone do not provide a good way to identify aftershocks, as there exist numerous examples of pairs of

events with nearly identical hypocentres, but separated in time by many months or years.

One simple resolution to this problem is to extend the definition of distance between events to include time separation as well as spatial separation. For example, suppose we define the 'space–time distance' d_{ST} between two events as

$$d_{ST} = \sqrt{d^2 + C^2 t^2}$$

where d is the usual geographic distance, t is the time separation, and C is a scaling constant to relate time and distance. Although the units of space–time distance are length, for example, km, they will be denoted by ST-km or space–time km to distinguish them from ordinary geographic distance units. Aftershocks can then be arbitrarily defined as the d-clusters for d_{ST} smaller than some fixed space–time distance D. It would be possible to define several other well-known metrics which would work nearly as well as d_{ST} for recognizing aftershocks, however this one is favoured for two reasons. First, this metric is the best approximation to the way the eye selects aftershock sequences on earthquake space–time plots, such as Fig. 8. Second, there is aesthetic appeal in the symmetry of the treatment of the temporal and the three spatial coordinates.

For d_{ST} to be useful in the identification of aftershocks which correspond to those selected subjectively by seismologists, C and D must correspond to our intuition about the possible relationships between events in space and time. For any particular region or catalogue, this will depend on the background rate of seismic activity, the density of events, and the magnitude cut-off used. In a companion paper submitted to this journal (Davis & Frohlich, in preparation) this method is applied to synthetic earthquake catalogues with a variety of background rates, and it is found that aftershocks can be successfully identified in several diverse catalogues using a scaling factor of $C = 1\,\text{km day}^{-1}$. Here, a value for C of $1\,\text{km day}^{-1}$ means that two events occurring in the same place 50 days apart are as closely related as two events 50 km apart occurring simultaneously.

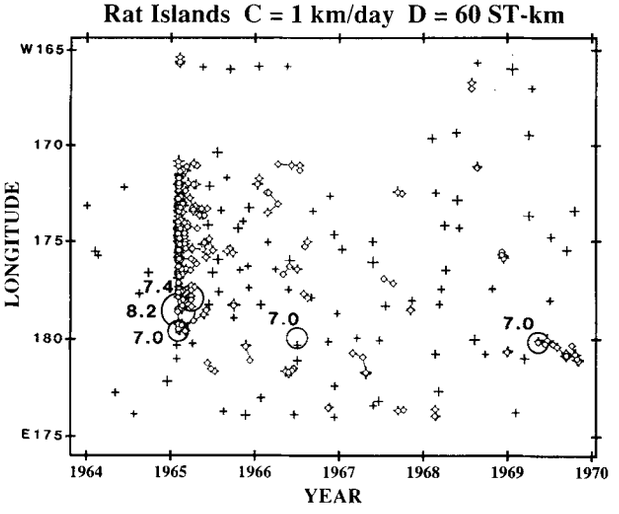

Figure 8. Space–time plot (longitude vs. time) of 432 earthquakes having m_b of 4.9 or greater and occurring along the western Aleutian trench as reported by the ISC between 1964 and 1970. Here earthquakes were linked using the space–time metric described in the text, relating time to distance by the parameter $C = 1.0\,\text{km day}^{-1}$. Lines denote links of length 60 ST-km or less. Open circles are linked earthquakes, and + 's are unlinked earthquakes. Note the Rat Islands earthquake 1965 February 4 ($M_S = 8.2$) and the Adreanof Islands earthquake of 1969 May 14 ($M_S = 7.0$) which are linked to large aftershock clusters. The Rat Island cluster includes 248 earthquakes occurring over a period of 153 days, while the Andreanof group has 21 aftershocks occurring over 176 days.

For example, if we consider ISC events in the western Aleutian trench having m_b of 4.9 or greater (Fig. 8), the authors' subjective selection of aftershock events agrees well with the occurrence of d-clusters when C is $1\,\text{km day}^{-1}$ and D is about 60 ST-km. For larger D (greater than 100 ST-km), the d-clusters contain events which seem unrelated, and for smaller D (less than 40 ST-km) some obvious examples of aftershocks are not linked. Using these parameters, most ISC earthquakes having m_b of 4.9 and greater are isolated events, and the majority of aftershock sequences occur as doublets or as small clusters of four or fewer aftershocks (Fig. 8). However, the Rat Islands earthquake of 1965 February 4, ($M_S = 8.2$) occurs near the beginning of a cluster of 248 events. Most of these earthquakes occur within a few days of the magnitude 8.2 earthquake, although some chains of aftershocks extend for months beyond this event.

In two companion papers (Davis & Frohlich, 1989 and in preparation), SLC as described here is used to investigate a variety of properties of aftershocks in the ISC catalogue, including Omori's law and the variation in aftershock occurrence in different tectonic regimes. In addition, in these companion studies SLC analysis is performed on simulated earthquake catalogues to determine what fraction of unrelated events SLC misidentifies as aftershocks, and what fraction of aftershocks SLC fails to register. SLC is also used to identify aftershock sequences in catalogues of events located by a local network of seismograph stations.

DISCUSSION AND CONCLUSIONS

In this paper it has been shown how SLC provides a quantitative framework for describing a number of earthquake phenomena which are most commonly described in qualitative terms. SLC is inherently descriptive, i.e. unlike a property such as earthquake moment which can be related to fault area and slip, the linkage and the statistics describing them have no direct physical significance. However, SLC does allow us to make meaningful comparisons of the properties of various earthquake nests, isolated events, aftershocks sequences, zones of seismic quiescence, etc. The present paper is the first in a series of papers where we apply SLC to study clustering properties of earthquake catalogues.

This paper has illustrated the application of SLC through the investigation of ISC events with m_b of 5.8 and greater occurring between 1964 January and 1986 February. For these data, the distribution of link lengths is smooth and continuous, suggesting that there is no natural gross scale distance which characterizes clustering within the data. However, the data for d-clusters for various values of d show that they closely resemble familiar tectonic groupings. That is, for d of about 500 km individual island arc groups

form distinct d-clusters, for d of about 200 km the d-clusters are large sub-groups within the major groups, and for d of 30 km or less the d-clusters are often earthquake nests or aftershock sequences.

The distribution of link lengths closely resembles that of synthetic events placed randomly along a circle with circumference of about 195 000 km. It is unclear whether or not there is any special physical significance in the similarity of the distribution of natural earthquakes links to that of the 1-D random events. Presumably it only signifies that earthquakes tend to occur along faults, rifts, and other long linear features, and that the total length of these features is about 4–5 times the Earth's circumference. However, both the biggest and smallest links are longer for the natural earthquake links than for the 1-D case. The longest links often join isolated event groups, such as earthquakes in Hawaii, to the Earth's major fault zones. Even though many of the shortest links join events in aftershock sequences, or events within groups of nuclear explosions, the links are longer than expected for the 1-D case because of location errors, and because fault zones typically have a finite width (Fig. 6).

In theory, the distribution of the lengths of links between events differs for event groups occurring within spaces of different dimensions. Moreover, for reasonably well-behaved processes, the peaks of the length distributions change in a predictable way over time (Appendix B). Thus for groups with many events, the length distribution provides a possible way to determine the dimension of the space, even if this space has a fractional dimension. However, to use SLC as a practical means for determination of dimensionality it would be necessary to undertake some analysis of synthetic catalogue data. This would allow the evaluation of how many events were necessary to determine the dimension accurately, and develop some schemes for recognizing composite distributions. We suggest that a likely place to search for variation in dimensionality is in catalogues from local networks, as the analysis presented here for the ISC data suggests that they are one-dimensional.

It is important to realize that the results for SLC or any clustering analysis depend strongly on the particular earthquake catalogue analysed. For example, for the ISC earthquakes having magnitude 5.8 and greater, it was found that the most intense nests of activity occurred at shallow depths in subduction zones (Table 3). Our analysis did not identify well-known nests at intermediate depths such as the Bucaramanga nest (Schneider, Pennington & Meyer 1987), because these earthquakes almost always have magnitudes smaller than 5.8. The Bucaramanga nest appears as a prominent cluster when SLC is applied to catalogues which include all earthquakes with magnitudes 5.0 and larger.

Furthermore, results obtained from SLC analysis may be influenced by the incompleteness or non-uniformity in earthquake catalogues, or by time-dependent variations in the determination of magnitudes or locations. These problems are especially severe when SLC is used to determine periods of quiescence (e.g. see Habermann 1988; Reasenberg & Matthews 1988). The principal advantage of SLC for this application is that it is fast and relatively objective. After some arbitrary choices about parameters such as cut-off level and scale distance are made, it can be applied entirely by a computer. For other applications, such as the identification of aftershocks, our preliminary analyses suggest that the SLC method is surprisingly reliable.

ACKNOWLEDGMENTS

We thank two anonymous reviewers for suggesting some significant improvements to an earlier version of this manuscript. The National Science Foundation provided funding for this research under grants EAR-8618406 and EAR-8843928. Support was obtained by S.D.D. from American Chemical Society grant PRF-18908-AC2.

REFERENCES

Amanzade, F. & Chatterje, S., 1984. Applications of clustering to exploration seismology, *Geoexploration,* **23,** 147–159.

Davis, S. D. & Frohlich, C., 1986. Application of single-link cluster analysis to the identification of foreshocks, aftershocks, and seismic gaps, *EOS,* **67,** 1119.

Day, W. H. E. & Edelsbrunner, H. 1984. Efficient algorithms for agglomerative hierarchical clustering methods, *J. Classification,* **1,** 7–24.

Eneva, M. & Pavlis, G. L., 1988. Application of pair analysis statistics to aftershocks of the 1984 Morgan Hill, California, earthquake, *J. geophys. Res.,* **93,** 9113–9125.

Frohlich, C., 1987. Aftershocks and temporal clustering of deep earthquakes, *J. geophys. Res.,* **92,** 13944–13956.

Frohlich, C. & Davis, S., 1985. Identification of aftershocks of deep earthquakes by a new ratios method, *Geophys. Res. Lett.,* **12,** 713–716.

Frohlich, C. & Davis, S. D., 1986. Single-link cluster analysis as a potential tool for evaluating spatial and temporal properties of earthquake catalogs, *EOS,* **67,** 1119.

Habermann, R. E., 1981. Precursory seismicity patterns: Stalking the mature seismic gap. In *Earthquake Prediction, an International Review,* pp. 29–42, eds Simpson, D. W. & Richards, P. G., American Geophysical Union, Washington, D.C.

Habermann, R. E., 1988. Precursory seismic quiescence: Past, present, and future, *Pure appl. Geophys.,* **126,** 279–318.

Habermann, R. E., McCann, W. R. & Nishenko, S. P., 1983. A gap is. . . , *Bull. seism. Soc. Am.,* **73,** 1485–1486.

Hartigan, J. A., 1975. *Clustering Algorithms,* John Wiley & Sons, New York.

Kagan, Y. Y., 1981a. Spatial distribution of earthquakes: the three-point moment function, *Geophys. J. R. astr. Soc.,* **67,** 697–717.

Kagan, Y. Y., 1981b. Spatial distribution of earthquakes: the four-point moment function, *Geophys. J. R. astr. Soc.,* **67,** 719–733.

Kagan, Y. & Knopoff L., 1978. Statistical study of the occurrence of shallow earthquakes, *Geophys. J. R. astr. Soc.,* **55,** 67–86.

Kagan, Y. Y. & Knopoff, L., 1980. Spatial distribution of earthquakes: the two-point correlation function, *Geophys. J. R. astr. Soc.,* **62,** 303–320.

Kagan, Y. & Knopoff, L., 1976. Statistical search for non-random features of the seismicity of strong earthquakes, *Phys. Earth planet. Int.,* **12,** 291–318.

Kagan, Y. Y. & Knopoff, L., 1981. Stochastic synthesis of earthquake catalogs. *J. geophys. Res.,* **86,** 2853–2862.

Kelleher, J., Sykes, L. & Oliver, J., 1973. Possible criteria for predicting earthquake locations and their application to major plate boundaries of the Pacific and the Caribbean, *J. geophys. Res.,* **78,** 2547–2585.

Ling, R. F., 1973. A probability theory of cluster analysis, *J. Am. Stat. Assoc.,* **68,** 159–169.

Mogi, K., 1979. Two kinds of seismic gaps, *Pure appl. Geophys.*, **117,** 1172–1186.

Ohtake, M., Matsumoto, T. & Latham, G. V., 1977. Seismicity gap near Oaxaca, southern Mexico as a probable precursor to a large earthquake, *Pure appl. Geophys.*, **115,** 375–385.

Prozorov, A. G. & Dziewonski, A. M., 1982. A method of studying variations in the clustering property of earthquakes: Application to the analysis of global seismicity, *J. geophys. Res.*, **87,** 2829–2839.

Reasenberg, P., 1985. Second-order moment of central California seismicity, 1969–1982, *J. geophys. Res.*, **90,** 5479–5495.

Reasenberg, P. A. & Matthews, M. V., 1988. Precursory seismic quiescence: A preliminary assessment of the hypothesis, *Pure appl. Geophys.*, **126,** 373–406.

Ripley, B. D., 1984. Spatial Statistics: Developments 1980–3. *Int. stat. Rev.* **52,** 141–150.

Schneider, J. F., Pennington, W. D. & Meyer, R. P., 1987. Microseismicity and focal mechanisms of intermediate-depth Bucaramanga nest, Colombia, *J. geophys. Res.*, **92,** 13913–13926.

Shlien, S. & Toksoz, M. N., 1970. A clustering model for earthquake occurrences, *Bull. seism. Soc. Am.*, **60,** 1765–1787, 1970.

Sneath, P. H. A. & Sokal, R. R., 1973. *Numerical Taxonomy*, W. H. Freeman, San Francisco.

Suzuki, Z. & Suzuki, K., 1965. On space distribution function of earthquakes, *Sci. Rep. Tohoku Univ., 5th Ser. (Geophys.)*, **17,** 9–23.

Suzuki, Z. & Suzuki, K., 1966. Change in spatial distribution of earthquakes against hypocentral depth, *Sci. Rep. Tohoku Univ., 5th Ser. (Geophys.)*, **17,** 159–168.

Utsu, T., 1972. Aftershocks and earthquake statistics (IV)—analyses of the distribution of earthquakes in magnitude, time, and space with special consideration to clustering characteristics of earthquake occurrence (2). *J. Fac. Sci., Hokkaido Univ., Ser. 7 (Geophys.)*, **4,** 1–42.

Vere-Jones, D., 1970. Stochastic models for earthquake occurrence, *J. R. Stat. Soc.*, **B32,** 1–62.

Wardlaw, R. L., Frohlich, C. & Davis, S. D., 1988. Analysis of seismic quiescence in the ISC catalog using single-link cluster analysis, *EOS*, **69,** 1299.

APPENDIX A

Algorithm to determine single-link clusters for N earthquakes

For a set S of N earthquakes numbered $i = 1, 2, \ldots, N$, we can form a square matrix $\mathbf{D}^1$ in which the element d_{ij}^1 in the ith row and the jth column corresponds to the distance between the ith and jth earthquake. Each element d_{ij}^1 is a possible link of length d_{ij}^1 between earthquake i and earthquake j. However, since there are only $N-1$ links, not all of the matrix elements d_{ij}^1 correspond to links. The problem is to describe an algorithm such that given $\mathbf{D}^1$, one can determine which of the d_{ij}^1 correspond to links.

First, note that each earthquake must be connected by a link to the earthquake closest to it in space. Thus if we search the ith row in $\mathbf{D}^1$ for the smallest off-diagonal element, this corresponds to the earthquake closest to i. Suppose it is element j; then (i, j, d_{ij}^1) describes a link. The closest earthquake to j may be earthquake i, or it may be a third earthquake k. If it is k then we have found two links corresponding to d_{ij}^1 and d_{jk}^1, otherwise our search for the closest events to i and j has produced only one link. Thus by this method at least $N/2$ links are found, and at most all $N-1$ elements (Fig. A1).

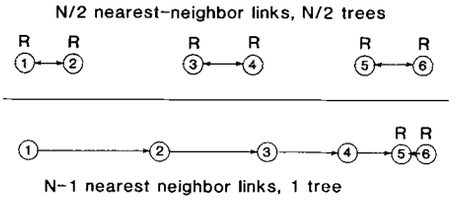

Figure A1. If N earthquakes (circles) in a set S are linked to their nearest neighbours, we will find at least $N/2$ links (top) or at most $N-1$ (bottom). The resulting linked groups are first-order trees. Root elements in trees are the two earthquakes which are each one another's nearest neighbour (marked R in diagram above). If the earthquakes in S are numbered $k = 1, 2, \ldots, N$, it is convenient to label non-overlapping subsets of S with the smallest value of k in the subset. In the top example, there are three trees labelled 1, 3 and 5. In the bottom example there is only one tree labelled 1.

If all the earthquakes are joined to their nearest neighbours as described, each linked group is called a *tree*. Note that the set of earthquakes in a tree is not the same as a d-cluster for any particular d, since the same tree may contain earthquakes for which d_{ij}^1 is very small and others where d_{ij}^1 is very large. If there are N earthquakes and M trees, to determine the remainder of the links it is necessary to determine the $M-1$ links in $\mathbf{D}^1$ which connect different trees. None of these remaining links that are required corresponds to any d_{ij}^1 where i and j are closest events, and none corresponds to earthquakes i and j in the same tree.

However, the remaining links can be found by a recursive procedure. Suppose a new matrix $\mathbf{D}^2$ is formed in which the element $d_{lm}^2 = d_{ij}^1$ is the minimum distance between the lth and mth tree, which occurs when the ith earthquake is linked to the jth earthquake. As before, the minimum off-diagonal element in row l of $\mathbf{D}^2$ will correspond to a link connecting tree l to its closest neighbouring tree. By finding all such 'second order links' and keeping track of the associated original events (i, j, d_{ij}^1), we will determine between $M/2$ and $M-1$ new links, which, when joined, form a new set of trees.

Repetition of this process and formation of the matrices $\mathbf{D}^3, \mathbf{D}^4$, etc., of distances between successively higher-order trees eventually leads to the point where all the trees are connected by $N-1$ links, as required.

When this process is complete, one reorders the links (i, j, d_{ij}^1) from largest to smallest d_{ij}^1, and the determination of the links for d-clustering of the N earthquakes is complete. To calculate using the algorithm as described here, it is necessary only to store $N(N-1)/2$ of the N^2 elements in $\mathbf{D}^1$, since the matrix is symmetric with $d_{ij}^1 = d_{ij}^1$ and the diagonal elements do not enter into the calculation.

When finding the $M-1$ tree-to-tree links, it is useful to label each tree with one of the elements. This can either be the element which appears earliest in the ordered set S, or one of the *root elements* (Fig. A1). The root elements are the two elements in every tree which are mutually closest to one another. At each step of the calculation one can keep track of the trees which are linked to one another by changing the labels of linked trees or tree groups so that linked groups have identical labels. Generally this is the smallest label value of events in the groups. Thus if a tree

with label 7 is linked to a tree with label 2, hereafter all elements in both trees would have label 2.

The authors have written a FORTRAN program which sets up the nearest-neighbour matrix $\mathbf{D}^1$, finds the nearest-neighbour links, and identifies and labels the first-order trees. It then sets up the tree nearest-neighbour matrix $\mathbf{D}^2$, and continues until all the links are found.

For more than a few thousand earthquakes it is not always feasible to determine linkage using the method described here. It requires computer memory which increases as N^2 and computation time which increases approximately as $N^2 \ln N$. Day & Edelsbrunner (1984) review this and various other approaches to efficiently determining linkage.

Fortunately, the distinctly non-random nature of the spatial distribution of earthquakes makes it possible to reduce the calculations significantly. In a typical situation, a group of earthquakes consists of a number k of visually identifiable, distinct sub-groups. In this case the links for the entire group are just the links as calculated for the individual sub-groups plus the $k-1$ elements linking the sub-groups. A distinct saving in computation time is achieved if the links for each of the sub-groups are calculated separately, and then the links joining the sub-groups found. It is possible to divide most sets of earthquakes into sub-groups visually, by inspection.

APPENDIX B

Some approximate results concerning link length distributions for randomly generated events in 1-, 2- and 3-dimensions

In this Appendix the approximate distributions of link lengths for events constrained to lie on manifolds of known dimensions are estimated. The efficacy of these approximate results is then illustrated in a comparison with computer-generated synthetics. It is also shown how these distributions will change as time passes and the number of events increases.

1-*Dimension*

Suppose a Poisson process, with rate λ_1 per unit length and time, places events along a linear region of length V_1. After time t, we expect that the number of events is $n_1(t) = \lambda_1 V_1 t$. If the 1-D region is straight, links must form between neighbouring events. Thus if we ignore the ends, the probability that a link longer than length r forms to the right of any given event is equal to the Poisson probability that no events occur within distance r to the right. The cumulative distribution function (CDF) $F_1(r)$ and probability density function (PDF) $f_1(r)$ for this are;

$$F_1(r) = 1 - \exp(-\lambda_1 r t),$$

$$f_1(r) = \lambda_1 t \exp(-\lambda_1 r t).$$

The PDF has no local maximum (Fig. B1a), but the median value for r [where $F_1(r_{1\text{med}}) = 0.5$] is;

$$r_{1\text{med}} = \frac{\ln 2}{\lambda_1 t}, \qquad f_1(r_{1\text{med}}) = \frac{\ln 2}{2 r_{1\text{med}}}.$$

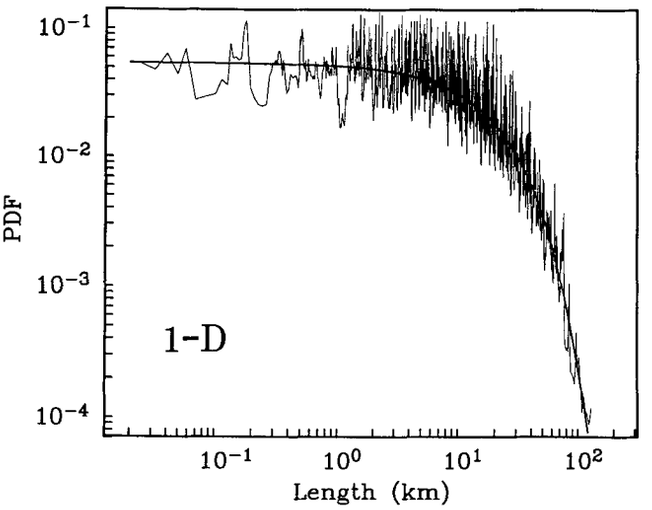

Figure B1a. PDF $f_1(r)$ of link lengths for events randomly distributed along a 1-D region. The smooth line is the analytical description derived in this Appendix, with $\lambda_1 t = 0.0544 = 2178/(2\pi R_{\text{Earth}})$. The jagged line is the synthetic result for 2178 events placed by the computer on the equator of a sphere having the Earth's radius R_{Earth}.

2-, 3- *and other dimensions*

For the 2-D and 3-D cases calculate the distribution of link lengths cannot be calculated exactly because not all links are nearest-neighbour links, or first-order links (see Appendix A). However, because at least 50 per cent of the links are always nearest-neighbour links (and in practice usually about 70 per cent), the time variation in link lengths will be estimated from a calculation of the properties of the nearest-neighbour links.

This calculation is analogous to the 1-D case, except here different formulas are used to find the area or volume near the given event. Poisson processes with rates λ_2 and λ_3 generate events over area V_2 and volume V_3, respectively. After time t the numbers of events are $n_2(t) = \lambda_2 V_2 t$, and $n_3(t) = \lambda_3 V_3 t$. We consider the probability that no event occurs within a distance r of a given event; the CDFs and PDFs are

$$F_2(r) = 1 - \exp(-\lambda_2 \pi r^2 t),$$

$$F_3(r) = 1 - \exp(-\lambda_3 \tfrac{4}{3}\pi r^3 t),$$

$$f_2(r) = 2\lambda_2 \pi r t \exp(-\lambda_2 \pi r^2 t),$$

$$f_3(r) = 4\lambda_3 \pi r^2 t \exp(-\lambda_3 \tfrac{4}{3}\pi r^3 t).$$

To find local maxima we differentiate $f_2(r)$ and $f_3(r)$ and set to zero, and find:

$$r_{2\max} = \left(\frac{1}{2\pi\lambda_2 t}\right)^{1/2}, \qquad f_2(r_{2\max}) = \frac{1}{r_{2\max}} \exp(-\tfrac{1}{2}),$$

$$r_{3\max} = \left(\frac{1}{2\pi\lambda_3 t}\right)^{1/3}, \qquad f_3(r_{3\max}) = \frac{2}{r_{3\max}} \exp(-\tfrac{2}{3}).$$

Thus as t increases, $r_{1\text{med}}$, $r_{2\max}$ and $r_{3\max}$ all decrease, while the peak values $f_1(r_{1\text{med}})$, $f_2(r_{2\max})$ and $f_3(r_{3\max})$ all increase. Figs B1(b) and B1(c) suggest that these approximate PDF's derived for nearest-neighbour links are quite similar to the computer-generated PDF's for SLC links, except that all

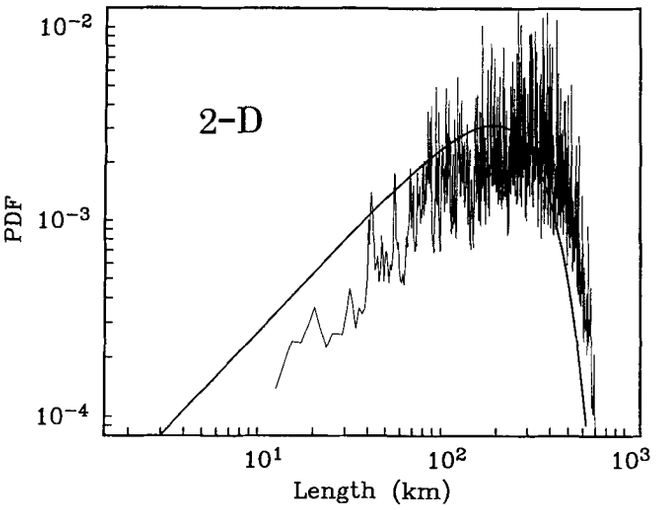

Figure B1b. PDF $f_2(r)$ of link lengths for events randomly distributed over a 2-D area. The smooth line is an approximate analytical description derived in this Appendix, with $\lambda_2 t = 4.27 \times 10^{-6} = 2178/(4\pi R^2_{\text{Earth}})$. The jagged line is the synthetic result for 2178 events placed by the computer on the surface of a sphere having the Earth's radius R_{Earth}.

links are slightly longer for the SLC PDF's. More generally, if the dimension is k, and if the rate of activity in k-space is $n_k(t) = c_k r^k t$, then the distribution of links will have a maximum at

$$r_{k\max} = \left(\frac{k-1}{kc_k t}\right)^{1/k}, \qquad f_k(r_{k\max}) = \frac{k-1}{r_{k\max}} \exp\left[-(k-1)/k\right].$$

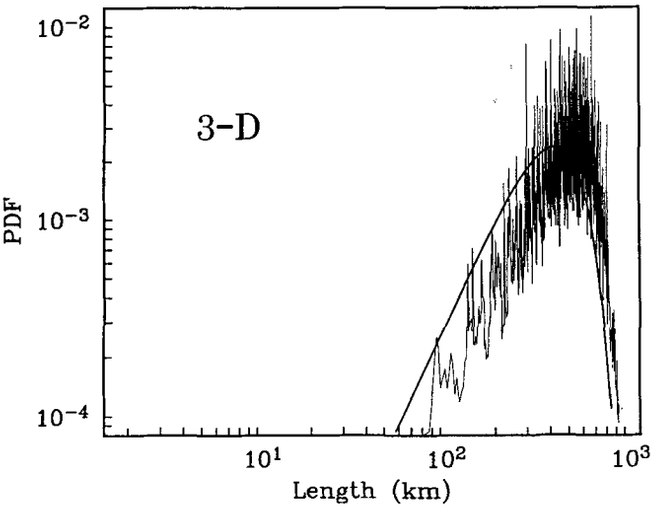

Figure B1c. PDF $f_3(r)$ of link lengths for events randomly distributed over a 3-D volume. The smooth straight line is an approximate analytical description derived in this Appendix, with $\lambda_3 t = 2.01 \times 10^{-9} = 2178/(\frac{4}{3}\pi R^3_{\text{Earth}})$. The jagged line is the synthetic result for 2178 events placed by the computer within a sphere having the Earth's radius R_{Earth}, as in Fig. 5(a).

Composite distributions

The results above can be used to approximate the distributions for events generated by combinations of 1-, 2- and 3-D processes. In particular, suppose that $w_1, w_2, \ldots, w_i$ are distribution weights such that $\sum_i w_i = 1$, and the w_i represent the relative contributions of distributions having PDF $f_{d_i}(r)$ to a composite process. Here d is the dimension of the individual PDF's, as derived above. In most cases, the majority of links will join events within a group rather than events in different groups. Thus, the PDF $f(r)$ for the composite process will be approximated by

$$f(r) = \sum_i w_i f_{d_i}(r).$$

Fig. B2 shows how such a composite process changes over time. Note that as time passes, the height of the peaks increases for the 2-D and 3-D sub-processes, and as predicted, $r_{d\max}$ becomes smaller. Over long enough time periods, the maximum for a composite distribution will be due to the sub-process with the smallest dimension, because $f_k(r_{k\max})$ is proportional to $t^{1/k}$.

If events occur only at a single point, all possible links will have length zero. If several such points exist in space, then the PDF f_0 will include a fixed component which represents the links joining discrete points in space, and a delta function at length zero for the zero-length links. For real earthquakes, the fixed component f_0 corresponds to the long links in Figs 5 and 6 which cross aseismic regions but join zones of seismic activity. Such zero-dimensional processes can also be superimposed onto other combinations of 1-, 2- and 3-D processes as described above.

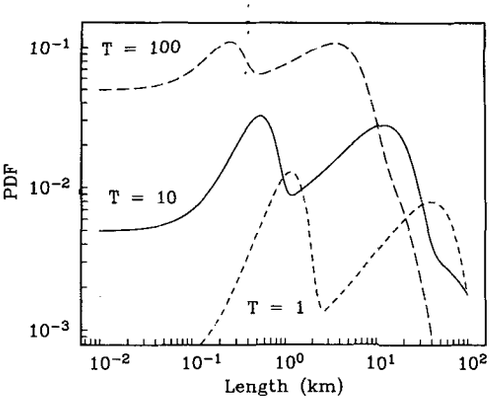

Figure B2. Variation over time for a composite PDF of a distribution having a 1-D component with $\lambda_1 = 10^{-3}$ and $w_1 = 0.493$, a 2-D component with $\lambda_2 = 10^{-4}$ and weight $w_2 = 0.493$ and a 3-D component with $\lambda_3 = 0.1$ and $w_3 = 0.014$. The rightmost dashed line is for $t = 1$, the solid line is for $t = 10$, and the leftmost dashed line is for $t = 100$.